Revisión de tema

Aproximación a la estadística inferencial: intervalos de confianza y pruebas de hipótesis

Approach to Inferential Statistics: Confidence Intervals and Hypothesis Testing

DOI: https://doi.org/10.11144/Javeriana.salud2.aeii

Eduardo José Pabón-Martínez [a]
Pontificia Universidad Javeriana , Colombia
ejpabonm@javeriana.edu.co
ORCID: https://orcid.org/0000-0001-6590-5933

Pedro José inzón-Castillo
Pontificia Universidad Javeriana, Colombia
ORCID: https://orcid.org/0000-0002-3645-6435

Valentina de Zubiría Sánchez
Pontificia Universidad Javeriana, Colombia
ORCID: https://orcid.org/0009-0005-7137-4485

María Alessandra Aroca
Pontificia Universidad Javeriana, Colombia
ORCID: https://orcid.org/0009-0004-0790-8852

Carlos Javier Rincón
Pontificia Universidad Javeriana, Colombia
ORCID: https://orcid.org/0000-0002-0477-1908

Recibido: 17 octubre 2024
Aceptado: 12 diciembre 2024

Resumen:

La estadística inferencial es una herramienta esencial en la investigación biomédica, cuyo uso ha aumentado significativamente en los últimos cincuenta años. Sin embargo, aún siguen existiendo deficiencias en la formación de los profesionales de la salud en los programas de pregrado y posgrado en esta área, reflejado en la mala interpretación de los resultados publicados en artículos científicos o al momento de realizar una investigación. Este artículo tiene como objetivos introducir la estadística inferencial, explicando conceptos clave como variables, probabilidad, parámetros y estadísticas, así como presentar los intervalos de confianza y las pruebas de hipótesis, enfatizando en la correcta interpretación de estos abordajes y sus elementos. La explicación conceptual se acompaña de un ejemplo que se construye y explica con figuras a lo largo del estudio, con el fin de lograr una mayor comprensión de estos y mostrar su uso en la investigación en salud.

Palabras clave: bioestadística, inferencia estadística, intervalos de confianza, pruebas de hipótesis.

Abstract:

Inferential statistics is an essential tool in biomedical research, the use of which has increased significantly in the last 50 years. However, there are still deficiencies in the training of health professionals in undergraduate and graduate programs in this area, reflected in the misinterpretation of the results published in scientific articles or when conducting research. This article aims to introduce inferential statistics, explaining key concepts such as variables, probability, parameters and statistics, as well as presenting confidence intervals and hypothesis tests, emphasizing the correct interpretation of these approaches and their elements. The conceptual explanation is accompanied by an example that is constructed and explained with figures throughout the study, in order to achieve a better understanding of these concepts and exemplify their use in health research.

Keywords: biostatistics, statistical inference, confidence intervals, hypothesis testing.

Notas de autor

[a]**Autor de correspondencia:** ejpabonm@javeriana.edu.co

Introducción

Un grupo de investigadores de medicina interna están interesados en conocer la edad promedio de los pacientes que ingresan por exacerbación de insuficiencia cardiaca sin diagnóstico previo a una unidad de agudos de un hospital de alta complejidad en Bogotá (Colombia). La hipótesis de los investigadores, basada en estudios en población europea, es que la edad promedio de ingreso es de 50 años. Para llevar a cabo el estudio, los investigadores seleccionaron una muestra aleatoria de 100 pacientes de un hospital de alta complejidad en Bogotá y, así, obtuvieron que la edad promedio fue de 52 años ($\underline{x} = 52$).

Los investigadores planean comunicar los resultados de la investigación, y ante ello plantean la siguiente pregunta: ¿la edad promedio de los pacientes que ingresan a una unidad de agudos del hospital de alta complejidad en Bogotá por exacerbación de insuficiencia cardiaca, sin diagnóstico previo, es igual a la reportada en los estudios previos (50 años)? Para responder la pregunta anterior, necesitamos la *estadística inferencial*, que permite obtener conclusiones sobre una población a partir de una muestra o subconjunto de unidades seleccionadas de esta.

Desde mediados del siglo XX, se ha observado un aumento progresivo en el uso de la estadística inferencial en la investigación biomédica (1) y ello ha llevado a que los profesionales de la salud deban comprender los conceptos relacionados con esta disciplina. Aun así, siguen existiendo deficiencias en la formación de los estudiantes en los programas de pregrado y posgrado (2), reflejadas en una mala interpretación de los resultados publicados en artículos científicos. Este artículo tiene como objetivos: 1) realizar una introducción a la estadística inferencial; 2) presentar conceptos relacionados como variable, probabilidad y distribución, y 3) presentar los intervalos de confianza y las pruebas de hipótesis, enfatizando en la correcta interpretación de estos abordajes y sus elementos, con el fin de precisar su uso. Las explicaciones se apoyan en figuras, fórmulas y la utilización de un ejemplo a lo largo del artículo.

Entendiendo lo básico: variable, probabilidad y distribución

Una *variable* es una característica que toma valores diferentes en cada observación hecha sobre un grupo de individuos o en el mismo individuo a lo largo del tiempo. Cuando los posibles valores que puede tomar una variable son resultado de un fenómeno aleatorio con una probabilidad de ocurrencia asociada, se denomina *variable aleatoria*. A partir de estas observaciones, se conoce la posibilidad de que se presente un cierto valor o conjunto de valores de una variable, para así obtener la proporción de las observaciones que toman el valor de interés del total de observaciones realizadas; estas proporciones se llaman *probabilidades empíricas* (3). Las probabilidades empíricas toman valores entre 0 y 1, y la suma de todas las probabilidades es igual a 1.

Dado que los valores que puede tomar una variable aleatoria tienen una probabilidad de ocurrencia, muestran un comportamiento que se representa en una *función de distribución (densidad) de probabilidades*, es decir, una expresión matemática que, al evaluarla para los valores de una variable, obtiene su probabilidad de ocurrencia (4). Dentro de las funciones de distribución de probabilidad, la más conocida es la distribución normal, pero existen muchas funciones, como la distribución binomial o la distribución X^2 (figura 1), con amplia aplicación en salud.

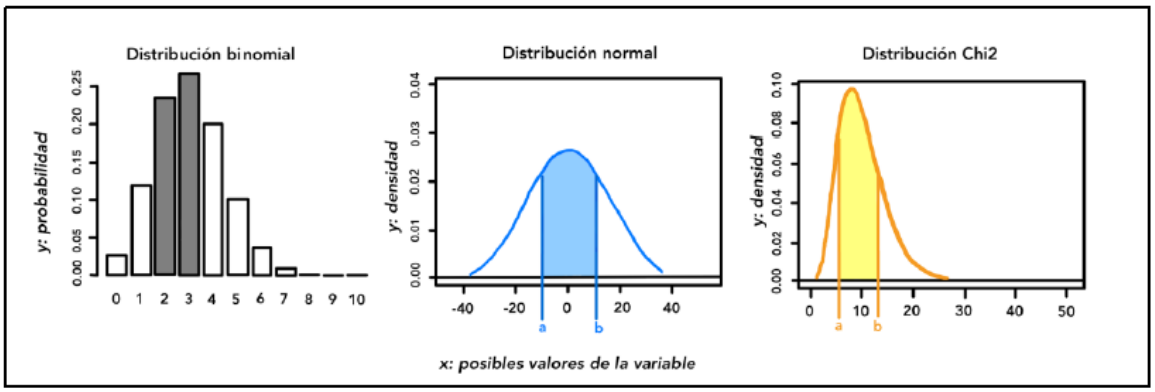

FIGURA 1
Ejemplos de funciones de distribución (densidad) de probabilidad

Conocer la función de distribución de una variable es crucial, porque permite responder a la pregunta: ¿cuál es la probabilidad de que los valores de la variable se encuentren entre el rango de valores A y B? Al conocer esta probabilidad, también podemos ejecutar el proceso inverso: determinar el valor o rango de valores que corresponde a una probabilidad específica.

Para nuestro ejemplo, la variable de interés es la edad. Dado que esta variable se comporta como una variable aleatoria, tiene una función de distribución, que para este caso asumimos es una distribución normal. Bajo esta distribución, podemos calcular la probabilidad de que la edad se encuentre en un rango entre 45 y 55 años. Asimismo, bajo esta distribución, también podemos responder preguntas, por ejemplo, a partir de qué valor de la variable se encuentra el 10 % de la población de mayor edad.

Construyendo la inferencia: de estadísticas a parámetros

Llamamos *población* al conjunto de individuos que comparten características comunes. De esta población, se obtiene un *parámetro* (θ), que es una medida calculada a partir de los valores de una variable en toda la población. El subconjunto de individuos o elementos de la población que utilizamos para realizar un estudio se denomina *muestra*. De esta muestra se obtiene una *estadística*, que es una medida calculada a partir de los valores observados en la variable de interés dentro de la muestra (5).

Para nuestro ejemplo, la población son los pacientes que ingresan a las unidades de agudos de los hospitales de alta complejidad en Bogotá por exacerbación de insuficiencia cardiaca, sin diagnóstico previo. Nuestro parámetro es la edad promedio de todos los pacientes con insuficiencia cardiaca que ingresan a estos hospitales. Nuestra muestra son los 100 pacientes del hospital de alta complejidad de Bogotá seleccionado, y la estadística es la edad promedio obtenida de nuestra muestra que, en este caso, corresponde a 52 años.

El objetivo de la *estadística inferencial* es representar a toda la población a partir de los valores observados en una muestra, es decir, se *busca conocer parámetros a partir de estadísticas.* Estas estadísticas se denominan *estimadores puntuales* ($\hat{\boldsymbol{\theta}}$)o *estadísticos de prueba* ($g(\hat{\boldsymbol{\theta}})$). Sin embargo, considere que de una población se pueden obtener muchas muestras (k) del mismo tamaño (n), donde cada muestra produciría una estadística diferente (figura 2). Por lo tanto, la estadística calculada a partir de una sola muestra no permite evaluar cómo varía la estimación respecto a otras posibles muestras y cómo se relacionan estas estadísticas con el parámetro de interés. Para abordar esta limitación, se han desarrollado los intervalos de confianza y las pruebas de hipótesis.

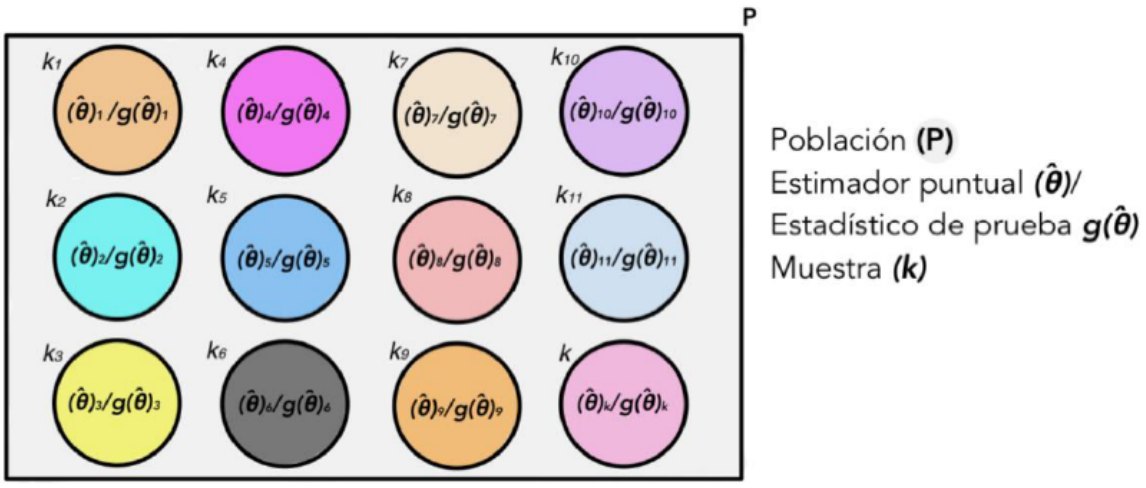

FIGURA 2

Representación de la población y muestras. En esta figura se representa una población y las posibles muestras que se pueden obtener. Es importante aclarar que si bien en aquí se muestran grupos en los cuales no se repiten participantes, en un escenario real las muestras no son mutuamente excluyentes

Intervalos de confianza: ¿qué son y para qué sirven?

Los intervalos de confianza son un rango de valores en el cual se encuentra, con un "nivel de confianza", el parámetro de interés (θ) En el marco de los intervalos de confianza, la estadística obtenida de una muestra específica se denomina *estimador puntual* ($\hat{\theta}$)(6). Imaginemos, nuevamente, que queremos repetir nuestro estudio muchas veces con diferentes muestras (k) del mismo tamaño (n) Cada muestra producirá un estimador puntual diferente. Si la media de los estimadores puntuales se aproximan al parámetro, a medida que aumenta el tamaño de la muestra, se define el estimador puntual como un estimador *insesgado* (figura 3).

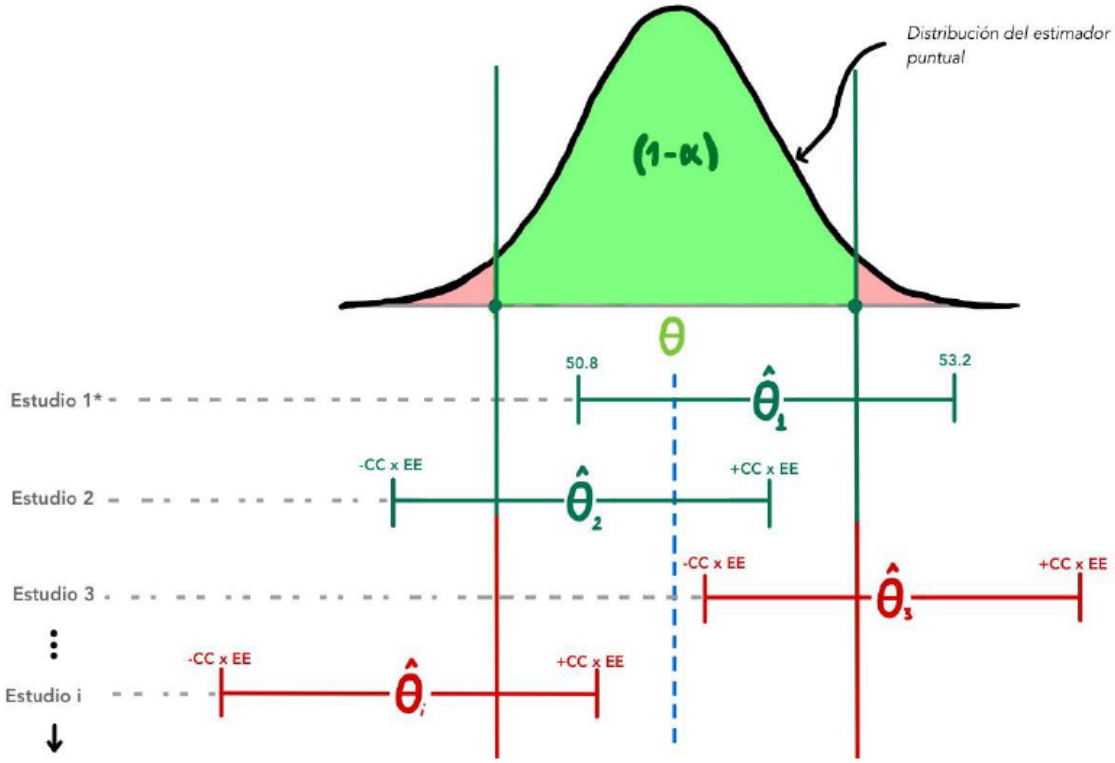

FIGURA 3
Representación de la inferencia por intervalos de confianza. Estudio 1, que corresponde a la representación de los intervalos de confianza del ejemplo

Dado que el estimador puntual ($\hat{\theta}$) puede tomar diferentes valores, se comporta como una *variable aleatoria* y, por lo tanto, sigue una *función de distribución de probabilidad*. Al conocer esta función, se calcula un rango de valores, o intervalos, que en la mayoría de las ocasiones incluirá el parámetro de interés (θ). Este intervalo se considera una "zona" alrededor de nuestro estimador puntual donde es muy probable que se encuentre el parámetro (7). La amplitud de este intervalo está determinada por la función $(1 - \alpha)$, la cual corresponde *al nivel de confianza* (véase figura 3), donde representa el *nivel de significancia*, cuyo valor es una cifra establecida por el investigador que, generalmente, es del 5% (0,05). De esta manera, al calcular este intervalo, afirmamos que en el $(1 - \alpha)$ % de las veces, el intervalo incluirá el parámetro de interés. Por ejemplo, si tenemos un nivel de confianza del 95% ($\alpha = 0,05$), en el 95 % de las ocasiones, el intervalo de confianza incluirá el valor del parámetro.

De forma general, un intervalo de confianza se compone por dos elementos adicionales al estimador puntual (8): el *error estándar (EE)*, que representa la variabilidad del estimador de una muestra a otra, y el *coeficiente de confiabilidad (CC)*, que corresponde al percentil[1] $(1 - \alpha/2)$ de la función de distribución de probabilidades del estimador y está relacionado con el nivel de confianza que deseamos para nuestro intervalo y depende del valor asignado a α.

Con estos elementos se calculan los *límites del intervalo de confianza* (fórmula 1) y se determina su precisión. Por ejemplo, si disminuimos el valor de α, aumenta el CC, y por ende, aumenta la amplitud del intervalo de confianza, reduciendo de esta manera su precisión. Por el contrario, si aumentamos el tamaño de la muestra, disminuimos la variabilidad del estimador puntual, ya que reducimos el EE. Así obtenemos menor amplitud del intervalo de confianza y una mayor precisión. Usualmente, en la investigación biomédica, como se señaló anteriormente, el valor de α suele ser el mismo ($\alpha = 0,05$), por lo que, para aumentar la precisión de los intervalos de confianza, se aumenta el tamaño muestral.

$$\hat{\theta} \pm CC \times EE$$ [Fórmula 1]

Para nuestro ejemplo, se sabe que la desviación estándar de la edad en la población es 6,2 años ($\sigma = 6,2$) y el estimador puntual es la media de la muestra ($\hat{\theta} = \bar{x} = 52$). Para un nivel de confianza del 95% y asumiendo que el estimador puntual tiene una función de distribución normal, el valor del CC corresponde a 1,96. El error estándar es $EE = \sigma/\sqrt{n}$, donde nuevamente α corresponde a la desviación estándar poblacional, y n, al tamaño de la muestra; por lo tanto, el $EE = 6,\frac{2}{10} = 0,62$. Al remplazar en la fórmula 1:

$$52 \pm 1,96 \times 0,62 = [50,8 - 53,2]$$

Por lo tanto, el intervalo de confianza corresponde a [*50,8 - 53,2*] (figura 3). Esto quiere decir que a un nivel de confianza del 95%, la edad promedio de los pacientes que ingresa por exacerbación de insuficiencia cardiaca en pacientes sin diagnóstico previo de insuficiencia cardiaca a un hospital de alta complejidad en Bogotá está entre 50,8 y 53,2 años.

Prueba de hipótesis: ¿qué son y para qué sirven?

Las pruebas de hipótesis son una herramienta similar a los intervalos de confianza que se usan en estadística inferencial para rechazar o aceptar una afirmación sobre un parámetro a partir de un valor obtenido en una muestra. El planteamiento de una hipótesis se basa en la formulación de la hipótesis nula ($H\theta$), que es una afirmación sobre un parámetro, y la hipótesis alternativa ($H\alpha$), qué es lo opuesto a $H0$ (9).

En nuestro ejemplo, la hipótesis de los investigadores de que la edad promedio de ingreso es de 50 años por estudios previos corresponde a la hipótesis nula (*H0: $\mu = 50$ años*), donde μ representa la edad promedio poblacional); mientras que la hipótesis alternativa sería lo contrario, es decir, que la edad promedio es diferente a 50 años (*Hα: $\mu \neq 50$ años*).

Ahora, si en la muestra la edad promedio fue de 52 años ($\bar{x} = 52$), ¿este valor es lo suficientemente distinto al reportado en la literatura para rechazar que la edad promedio poblacional sea 50 años? Con la finalidad de responder esta pregunta se crearon las pruebas de hipótesis.

Como ya se ha mencionado, de una población se pueden obtener infinitas muestras del mismo tamaño, donde cada una genera una estadística, que en el marco de las pruebas de hipótesis se denomina *estadístico de prueba* ($g(\hat{\theta})$).Teniendo en cuenta nuevamente que cada estadístico de prueba es una cantidad que varía según la muestra, se comporta como una variable aleatoria y, por ende, tendrá una función de distribución de probabilidades.

Ahora, si se asume que la *H0* es cierta, el estadístico de prueba seguirá una función de la distribución de probabilidades, donde se establecen puntos de referencia (*valores críticos*), para delimitar un área *de no rechazo* en la cual se encuentran los valores del estadístico de prueba sin evidencia suficiente para rechazar *H0*. Su complemento, que corresponde a la *zona de rechazo*, es el área donde se encuentran los valores del estadístico de prueba que tienen evidencia necesaria para rechazar *H0* en favor de la *Hα* (figura 4a). Estos valores críticos, que definen ambas áreas, están determinados por el valor de α, que en el contexto de la prueba de hipótesis se denomina *error tipo I o nivel de significancia*, y corresponde a la probabilidad de rechazar la hipótesis nula (H0) cuando esta es verdadera. El valor de α es predefinido por el investigador y, generalmente, se establece en un 5% (0,05). Ahora, ($1 - \alpha$), que es su complemento, corresponde a la probabilidad de no rechazar *H0* cuando es cierta, y se denomina *nivel de confianza*. Este nivel de confianza es el *criterio utilizado para determinar qué tan distante considero suficiente para rechazar H0* (10,11) (figura 4a).

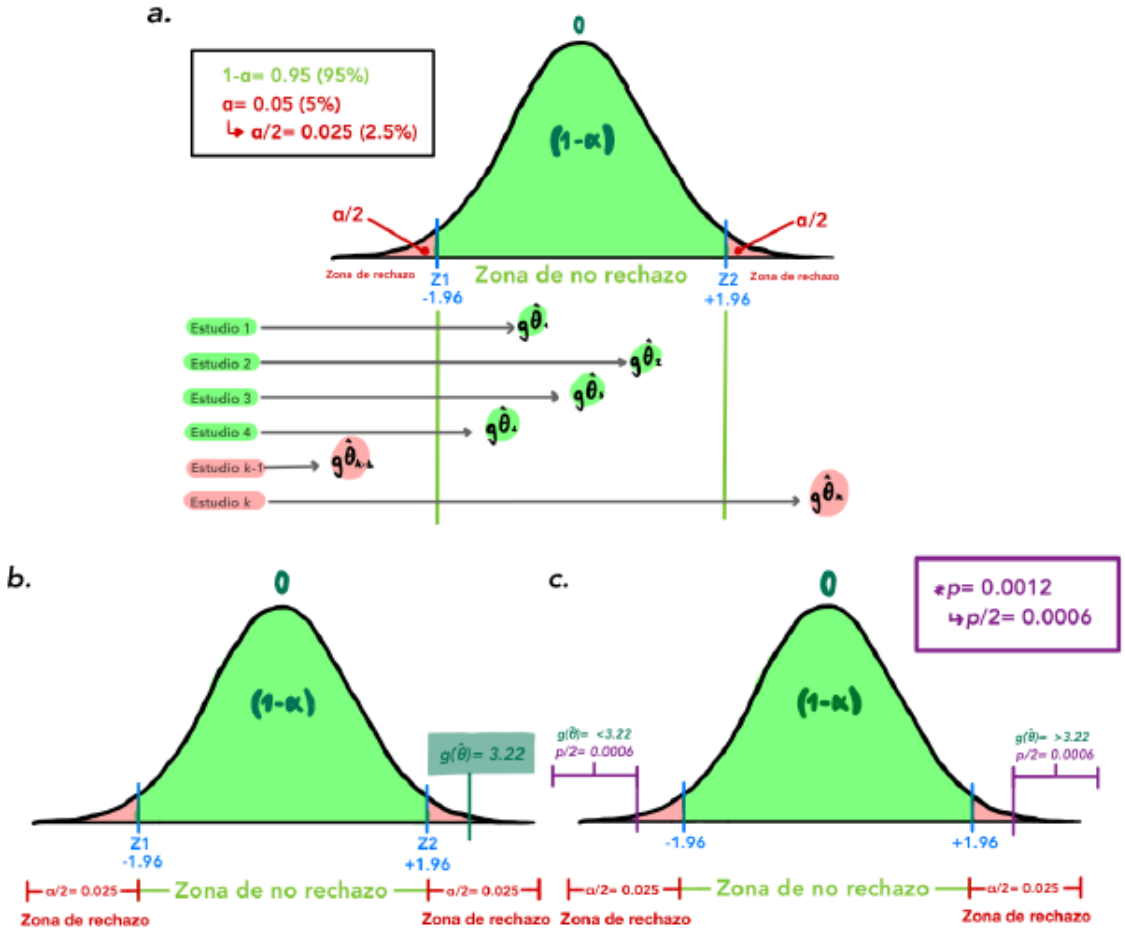

FIGURA 4

a) Representación de la inferencia por pruebas de hipótesis. **b)** Representación del estadístico de prueba [Zg(θ)] en la distribución normal. **c)** Representación del valor de *p* para la distribución normal

Adicionalmente, en caso de que la *H0* sea falsa, la distribución del estadístico de prueba es diferente y se describen dos conceptos adicionales.

- → β es la probabilidad de no rechazar *H0* cuando esta sea falsa. Este valor también se denomina *error tipo II*.
- → 1 - β es la probabilidad de rechazar *H*θ en favor de *Ha* cuando *H0* sea falsa, también denominado *poder* (9).

Al seguir este proceso, se desea ubicar el estadístico de prueba en la distribución de probabilidades para definir si se encuentra en la zona de rechazo o de no rechazo y tomar una decisión en favor o en contra de la hipótesis nula.

Para nuestro ejemplo, sabemos que $(\underline{x} = 52)$. años, la *H0* es que la media poblacional es 50 años (debido a hallazgos en otros estudios), y la desviación estándar poblacional es conocida y corresponde a 6,2 años. El cálculo del estadístico de prueba es igual a $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{52-50}{6.2/\sqrt{100}} = 3,226$ (11). Se ubica el valor de $g(\hat{\theta})$ en la distribución normal estándar, que bajo un $\alpha = 0,05$, los valores críticos corresponden a −1,96 y 1,96, por lo que el estadístico de prueba se encuentra en la zona de rechazo. Esto quiere decir que la edad promedio del paciente con insuficiencia cardiaca que ingresa a la unidad de agudo de un hospital de alta complejidad es diferente a 50 años ($H\alpha : \mu \neq 50$ *años*) y se puede rechazar la *H0* en favor de la *Hα*. En la figura 4b se grafica la ubicación de $g(\hat{\theta})$ en la distribución normal estándar.

Valor de p: ¿cómo se interpreta?

Adicionalmente, como parte de la prueba de hipótesis, aparece el concepto de *valor de p*, que se define como la probabilidad de que el estadístico de prueba se encuentre en los valores más extremos de la distribución a partir del $g(\hat{\theta})$ obteniendo en la muestra (12).

- → Si el valor de p es *mayor o igual* al nivel de significancia (α), el $g(\hat{\theta})$ estará en la zona de no rechazo y se concluirá que no existe evidencia significativa para rechazar *H0*.
- → Si el valor de p es *menor* al nivel de significancia (α), el $g(\hat{\theta})$ estará en la zona de rechazo y se concluirá que existe evidencia significativa para rechazar *H0* en favor de (13).
- → Para nuestro ejemplo con $g(\hat{\theta})= 3{,}22$, la probabilidad de que el $g(\hat{\theta})$ se encuentre en los valores más extremos al valor del estadístico de prueba observado es igual a la probabilidad de que el estadístico de prueba sea mayor a 3,22 o sea menor a $-3{,}22$. Bajo la distribución normal, el valor de p es igual $p = Prob(g(\hat{\theta}) \leq -3{,}22) + Prob(g(\hat{\theta}) > 3{,}22) = 2 * 0{,}0006 = 0{,}0012$.

Finalmente, se puede decir que el valor de p es menor que el nivel de significancia ($\alpha = 0{,}05$), por lo que se encuentra en la zona de rechazo y hay evidencia significativa para rechazar *H0* en favor de *Ha*. Es decir, la probabilidad de obtener un resultado tan extremo como el valor del estadístico de prueba observado es muy pequeño y ello sugiere que el promedio poblacional de la edad no es igual a 50 años. En la figura 4c se graficó el valor *de p* de nuestro ejemplo.

Conclusiones

En el presente artículo se exploró la estadística inferencial, desde la comprensión de conceptos básicos como variable, probabilidad y distribución de probabilidad, hasta el desarrollo de pruebas de hipótesis, los intervalos de confianza y el valor de p. A lo largo de la construcción del concepto se proporcionó un ejemplo que aterriza las definiciones, explicando además la importancia real y uso de cada uno de estos elementos. Así, se logra una comprensión profunda que trasciende el acto mecánico de aplicar fórmulas sin entender su propósito. De esta manera, se destaca la importancia real de la estadística inferencial, su correcta interpretación y su aplicación adecuada en el contexto de la investigación biomédica.

Agradecimientos

Al Semillero de Bioestadística de la Pontificia Universidad Javeriana, por su compromiso con la educación de bioestadística en los profesionales de la salud.

Referencias

1. Altman D, Machin D, Bryant T, Gardner M. Statistics with confidence. 2.ª ed. Bristol: BMJ Books; 2000.
2. Fardales-Macías V, Fábregas-Tejeda J, Carrazana-Rodríguez R. Insuficiencias en la preparación estadística del estudiante de medicina. Medisur [internet]. 2017;15(4):493-508. Disponible en: https://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=76062
3. Ruiz Morales A, Gómez-Restrepo C. Medidas de frecuencia, de asociación y de impacto. En: Epidemiología clínica. 2.ª ed. Bogotá: Editorial Médica Panamericana; 2015. p. 113-28.
4. Dawson B, Trapo RG. Basic and clinical biostatistics. 3.ª ed. Nueva York: McGraw-Hill; 2001.

5. Diana MF. Conceptos básicos sobre bioestadística descriptiva y bioestadística inferencial. Rev Argentina Anestesiol. 2006;64(6):241-51.

6. Daniel W. Bioestadística base para el análisis de las ciencias de la salud. México: Limusa; 2002.

7. Ochoa Sangrador C, Molina Arias M, Ortega Páez E. Inferencia estadística: probabilidad, variables aleatorias y distribuciones de probabilidad. Evid Pediatr. 2019;15:27.

8. Castañeda J, Fabián Gil J. Una mirada a los intervalos de confianza en investigación. Rev Colomb Psiquiatr [internet]. 2004;33(2):193-201. Disponible en: https://www.redalyc.org/pdf/806/80633208.pdf

9. Altamirano AM, Villa Romero AR. Bases para el razonamiento en estadística inferencial. En: Epidemiología y estadística en salud pública. México D. F.: McGraw-Hill Interamericana; 2011.

10. Payne TH. Statistical interpretation of data for clinical decision making. En: Goldman-Cecil medicine. 27.ª ed. Philadelphia, PA: Elsevier; 2024.

11. Dagnino J. La distribución normal. Rev Chil Anestesiol [internet]. 2014;43(2):116-21. Disponible en: https://revistachilenadeanestesia.cl/PII/revchilanestv43n02.08.pdf

12. Ballard SB, Blazes DL. Epidemiología aplicada para el médico de enfermedades infecciosas. En: Epidemiología de las enfermedades infecciosas. 9.ª ed. Barcelona: Elsevier; 2021.

13. White SE. Probabilidad y temas relacionados para hacer inferencias sobre los datos. En: Bioestadística básica y clínica. 5.ª ed. Ciudad de México: McGraw-Hill Education; 2021.

Notas

1 Un percentil (k) corresponde al valor de la variable donde el k % de las observaciones son menores o iguales a este valor.

Cómo citar: : Pabón-Martínez EJ, Pinzón-Castillo PJ, De Zubiría Sánchez V, Aroca MA, Rincón CJ, Aproximación a la estadística inferencial: intervalos de confianza y pruebas de hipótesis. Salud Javeriana. 2025; 2. https://doi.org/10.11144/Javeriana.salud2.aeii